

CSC2231 – Internet Systems and Services

Paper Review #1 – A Case for NOW (Networks of Workstations)

Name: Alex Wun
Student #: 990924737
Date: Sep. 14, 05

Anderson and his colleagues argue that economies of scale are the fundamental driving force for innovation and technological advance in computers. Mainframe and supercomputer technologies evolve slowly since there is less return on research investments due to low volumes of sale. In contrast, workstations and desktops provide the best performance/price ratio due to much higher sales volumes and heavy research investments. Because of this, performance and resource intensive jobs should be run on Networks of Workstations rather than traditional dedicated supercomputers.

Networked workstations effectively become a dynamic “pool” of resources that offer aggregated processing power, main memory, and secondary storage for large jobs. Networked workstations can provide parallel-processing power capable of matching (and in some cases, exceeding) the capabilities of traditional supercomputers. Additionally, a vast quantity of main memory is made available for file caching or use as virtual memory over the network. Finally, the potential for a flexible and dynamic network facilitated RAID is also made possible.

The key strength of this approach is that all these aggregated resources are taken from regular workstations and not machines designed specifically for large jobs. This allows organizations to leverage inexpensive commercial workstations that are also used for normal everyday tasks. A significant challenge addressed by the paper is being able to allocate idle resources intelligently for large jobs without affecting the small, interactive jobs that run locally on each individual workstation.

However, the success of this approach depends heavily on the infrastructure being used. Since the network is acting as an I/O bus, low-latency networks are a necessity. But even with such a network, the latencies of an actual I/O bus cannot be matched – which is a significant limitation. Similarly, each workstation must meet minimum technology levels in order to be useful, which means that only high-end workstations are viable. Otherwise, the overhead of workstation management and network communications will outweigh the benefits of having cheaply aggregated resources available. Finally, although the paper asserts that workstations remain largely idle during the day, it is not guaranteed (nor is this usage pattern necessarily representative of all organizations). Workstation usage fluctuates in much the same way that electricity usage has peaks and valleys. This means that resource availability for large jobs is not guaranteed, implying that job execution times are also not guaranteed (unlike the dedicated supercomputer scenario). The NOW can only provide “best effort” services.

In conclusion, there is clear evidence that the proposed model of pooling network resources is at the forefront of distributed computing developments. Some notable examples include IBM’s focus on GRID computing, the development of Storage Area Networks (e.g. iSCSI – any network host adhering to the iSCSI protocol can act as secondary storage), and P2P filesharing.